

This article was downloaded by: [López-Fidalgo, J.]

On: 27 October 2010

Access details: Access Details: [subscription number 928672332]

Publisher Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of Applied Statistics

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713428038>

### Design issues for population growth models

J. López Fidalgo<sup>a</sup>; I. M. Ortiz Rodríguez<sup>b</sup>; Weng Kee Wong<sup>c</sup>

<sup>a</sup> Departamento de Matemáticas, Universidad de Castilla la Mancha, Castilla La Mancha, Spain <sup>b</sup>

Departamento de Estadística y Matemática Aplicada, Universidad de Almería, Almería, Spain <sup>c</sup>

Department of Biostatistics, University of California at Los Angeles, Los Angeles, USA

First published on: 27 October 2010

**To cite this Article** Fidalgo, J. López, Rodríguez, I. M. Ortiz and Wong, Weng Kee (2010) 'Design issues for population growth models', Journal of Applied Statistics, First published on: 27 October 2010 (iFirst)

**To link to this Article:** DOI: 10.1080/02664760903521419

**URL:** <http://dx.doi.org/10.1080/02664760903521419>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Design issues for population growth models

J. López Fidalgo<sup>a</sup>, I.M. Ortiz Rodríguez<sup>b</sup> and Weng Kee Wong<sup>c\*</sup>

<sup>a</sup>*Departamento de Matemáticas, Universidad de Castilla la Mancha, Castilla La Mancha, Spain;*

<sup>b</sup>*Departamento de Estadística y Matemática Aplicada, Universidad de Almería, Almería, Spain;*

<sup>c</sup>*Department of Biostatistics, University of California at Los Angeles, Los Angeles, 90095, USA*

(Received 15 March 2009; final version received 10 November 2009)

We briefly review and discuss design issues for population growth and decline models. We then use a flexible growth and decline model as an illustrative example and apply optimal design theory to find optimal sampling times for estimating model parameters, specific parameters and interesting functions of the model parameters for the model with two real applications. Robustness properties of the optimal designs are investigated when nominal values or the model is mis-specified, and also under a different optimality criterion. To facilitate use of optimal design ideas in practice, we also introduce a website for generating a variety of optimal designs for popular models from different disciplines.

**Keywords:** approximate design; c-optimal design; D-optimal design; equivalence theorem

## 1. Introduction

Nonlinear models are ubiquitous in all biological areas of science. Lindsey [15] provided an eclectic collection of nonlinear models used in areas such as epidemiology, ecology, pharmaceutical and biomedical areas. The work of Motulsky and Christopoulos [19] is another exemplary monograph that catalogues applications of nonlinear models in various disciplines. In toxicology, risk assessment models are frequently used to study dose–response relationship. Our interest is in growth models that appear in numerous disciplines, such as animal science, plant growth and cell proliferation. For example, Jolicoeur *et al.* [10] reviewed and proposed new models for longitudinal growth of human stature. More general reviews of growth models include Fujii [6] and Karkach [12]. Fujii [6] reviewed the numerous and ever-growing growth models in auxology aimed at studying human time-series variations. Karkach [12] also provided an in-depth review of models for individual growth rates for various species in animals, plants and human. The author noted that one reason for the abundance of research in this area is that different parts of organisms, such as in plants, can grow with different patterns. There are many additional growth models proposed and they include Jolicoeur *et al.* [11], Jolicoeur and Pontier [9], Jolicoeur *et al.* [10], and, Martin *et al.* [17], just to name a few. It should be noted that longitudinal growth of stature of individuals (i.e. body growth) and growth in numbers of individuals in a population are different

---

\*Corresponding author. Email: wkwong@ucla.edu

processes. The former case is essentially a biologically continuous process, but population growth model is a continuous approximation to a discrete process that models increase of individuals in large populations.

Design issues for experiments in the biological sciences seem to be generally given short shrift. Many scientific papers do not justify the design employed for their studies even though the cost is readily acknowledged. In some studies, the study cost can be prohibitive because it requires a long time to obtain an appropriate number of observations. Statistical considerations can be incorporated into the construction of designs that result in improved inference without increasing cost. For example, López-Fidalgo and Wong [16] applied optimal design theory and constructed a variety of efficient designs for the widely used Michaelis–Menten model. Specifically, they found optimal designs for the Michaelis–Menten model for estimating one or two parameters and/or when there are competing interest in the two parameters. The authors showed that the optimal designs required fewer design points and offered substantial saving over designs implemented in the studies, including the popular uniform designs that take equal number of observations over a set of equally spaced points.

In this paper, we are interested in addressing design issues for models that study growth and decline of a population. Like in many other applied disciplines, the bulk of the papers in the growth literature addressed modeling and estimation issues but not design issues. Our goal is to choose optimal sampling times and how many to select at each time point. To fix ideas, we choose the model proposed by Jolicoeur and Pontier [9] as an illustrative case to find an optimal design for estimating model parameters and interesting functions of model parameters. The mathematical model is

$$E(Y_t) = \frac{1}{C_1 e^{t/D_1} + C_2 e^{-t/D_2}}, \quad t \in [0, \infty), \tag{1}$$

where  $C_1, C_2, D_1, D_2 \geq 0$ ,  $t$  is the time and  $Y_t$  the size or the proportion of the population at time  $t$ . The parameter  $D_1$  is responsible for the growth of the population,  $D_2$  measures the decline rate and  $C_1$  and  $C_2$  simultaneously control the relative rates of increase, decrease and height of the curve, that is the population size. This model is parsimonious and is used in ecology for studying population growth and decline in animals, or for studying the growth and decline of molecular or cellular populations within individual organisms.

Model (1) reduces to the logistic growth curve when  $(D_1 \rightarrow \infty)$  and to the logistic decline curve when  $(D_2 \rightarrow \infty)$ . It also reduces to the growth exponential model when  $(C_1 = 0)$  and to the decline exponential model when  $(C_2 = 0)$ . These limiting forms for the proposed model are given explicitly in Table 1 in [9]. Figure 1 shows the illustrative behaviors of these four cases subsumed in model (1).

Table 1. Locally  $c$ -optimal designs for the *P. caudatum* example with sampling times at 1.27, 3.38, 5.85 and 16, and locally  $c$ -optimal designs for the *M. variegata* example with sampling times at 8.15, 16, 24.57 and 55.50.

	Types of optimal designs	Weights
<i>P. caudatum</i>	$c_{C_1}$ -optimal design	0.07, 0.06, 0.37, 0.50
	$c_{C_2}$ -optimal design	0.74, 0.14, 0.07, 0.05
	$c_{D_1}$ -optimal design	0.04, 0.03, 0.20, 0.73
	$c_{D_2}$ -optimal design	0.50, 0.24, 0.15, 0.11
<i>M. variegata</i>	$c_{C_1}$ -optimal design	0.08, 0.06, 0.33, 0.53
	$c_{C_2}$ -optimal design	0.67, 0.16, 0.10, 0.07
	$c_{D_1}$ -optimal design	0.05, 0.04, 0.20, 0.71
	$c_{D_2}$ -optimal design	0.50, 0.23, 0.15, 0.12

Notes: All locally  $c$ -optimal designs were found using nominal values from Jolicoeur and Pontier (1989).

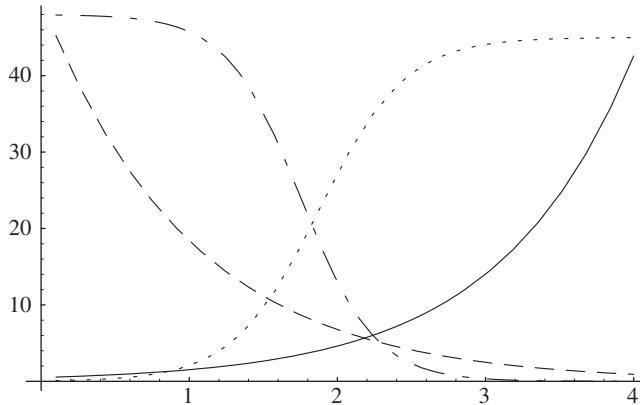


Figure 1. Comparison of the behavior of four particular cases of model (1): exponential growth (—), logistic growth (· · ·), exponential decline (---) and logistic decline (— · —).

Our purpose is to construct efficient designs for estimating model parameters ( $D$ -optimality) or specific parameters or linear combination of model parameters ( $c$ -optimality).  $D$ -optimality is used when interest is in estimating all model parameters. A  $D$ -optimal design minimizes the volume of the confidence ellipsoid of the parameters, thereby ensuring that we have the most precise estimates for the parameters. Sometimes, a function of the model parameters is of interest. In this case,  $c$ -optimality is the design criterion typically used for this purpose and its name comes from the optimal design literature. As an example, consider model (1) and we want to estimate the time to maximal population growth or decline in model (1). In this case, time to maximal growth may be found by first differentiating the function in (1) with respect to  $t$  and solving the resulting equation. This means time to maximal growth or decline is a function of the model parameters and our interest is to estimate this function accurately. A  $c$ -optimal design fulfills this purpose by minimizing the variance of the estimate of this function. Such optimal designs can offer substantial savings in time, labor and cost and at the same time produce the most precise estimates by careful choice of the design. In our present context, such an optimal design will provide guidance on the number of sampling time points, the time points and how many to sample at each of these optimal sampling time points to get the most precise estimate once we are given a fixed amount of resources.

In the next section, we describe the statistical set up and briefly review optimal design theory. Because of the complexity of the model, we only present numerically optimal designs, i.e. no analytical formulae for the optimal designs. The optimal designs provide valuable guidance for researchers in terms of cost savings and also a meaningful way for evaluating alternative designs. In Section 3, we present two illustrative examples that will be used to construct locally  $D$ - and  $c$ -optimal designs for model (1). Section 4 discusses properties of locally  $D$ - and  $c$ -optimal designs for model (1) and their robustness properties to model assumptions. We conclude with an appendix that provides some technical justifications for our results.

## 2. Background

We assume that the total number of observations is  $N$  and is pre-determined by experimental cost constraints. An experimental design consists of a collection of points  $t_1, t_2, \dots, t_N$ , in a given interval  $\mathcal{T}$ . A convenient way to understand designs is to treat a design as a collection of different points from  $\mathcal{T}$ , along with the proportion of the  $N$  observations allocated at these points. These are the so-called approximate designs and they are essentially probability measures on

$\mathcal{T}$ . Symbolically, we denote an  $n$ -point approximate design by  $\xi$ , its design points by  $t_1, \dots, t_n$  and its weight distribution over these points by  $\xi(t_1), \dots, \xi(t_n)$ . In practice, such a design takes approximately  $N\xi(t_i)$  observations at  $t_i, i = 1, \dots, n$  subject to  $N\xi(t_1) + \dots + N\xi(t_n) = N$ . As an example, suppose we have a model to represent the growth of deer population in a region over time and we want to estimate the peak population in the region over a time period  $\mathcal{T}$ . Assume further that we have resources to sample  $N = 100$  deers from the region during this time period. The research question is how to sample these 100 deers over the time period  $\mathcal{T}$ ? The  $c$ -optimal design  $\xi$  tells us to take a sample at  $n$  time points and roughly  $N\xi(t_i)$  deers at time  $t_i, i = 1, \dots, n$  subject to the restriction that  $N\xi(t_1) + \dots + N\xi(t_n) = N$ .

Approximate designs have increasingly proven to be an effective tool for studying design issues [1,23,24]. This approach allows us to find optimal designs for many problems that were otherwise impossible to determine and more importantly, there are algorithms for finding a variety of optimal designs. Another advantage of this approach is that the theory is well developed; in particular, it allows us to search for the optimal design within approximate designs with a small number of support points. This simplifies and speeds up the search for the optimal design and in many cases enables us to find the optimal design analytically as well.

Throughout we suppose our mean outcome, i.e. average population growth can be adequately described by a regression model with time  $t$  as the single variable in the model. All random errors in the model are assumed to normally and independently distributed, each with mean 0 and constant variance. The mean outcome denoted by  $\eta(t, \theta)$  is assumed to be known apart from the vector of model parameters  $\theta$ . The design points are to be selected from a given time interval  $\mathcal{T}$  and they represent sampling time points to observe the population growth.

Let  $f(t, \theta) = \partial \eta(t, \theta) / \partial \theta$ . Under the normality assumption, the Fisher information matrix of a design  $\xi$  is given by

$$M(\xi, \theta) = \sum_{t \in \mathcal{T}} f(t, \theta) f^T(t, \theta) \xi(t), \quad (2)$$

apart from an unimportant multiplicative constant. The inverse of this matrix is asymptotically proportional to the variance–covariance matrix of the estimated parameters. For this reason, many design optimality criteria are formulated in terms of the information matrix. We sometimes denote  $M(\xi, \theta)$  simply by  $M(\xi)$  when there is no confusion.

Unlike linear models, the information matrix of a design for a nonlinear model depends on the model parameters. This means that our optimal design depends on the parameters that we are trying to estimate! Frequently, as a first step, we assume nominal values of the parameters are available either from similar studies or from pilot studies. Nominal values represent the best guesses for the true values of the parameters based on existing knowledge. Local optimal designs are then constructed based on nominal values [2]. In practice, after the locally optimal design is determined from the nominal values, we use data from the design to estimate the parameters again. This iterative procedure is repeated and usually converges to some stable values for the parameters. This roundabout approach may seem awkward but frequently is a first and effective step in designing for a nonlinear model.

We consider only locally optimal approximate designs and focus on two optimality criteria in this work:  $D$ -optimality for estimating the model parameters  $\theta$  and  $c$ -optimality for estimating a user-selected function of the model parameters. The  $D$ -optimality criterion minimizes the first-order approximation of the volume of the confidence ellipsoid of  $\theta$  and is given by  $\Phi_D[M(\xi)] = \det M^{-1/m}(\xi)$ , where  $m$  is the number of parameters in the model.  $D$ -optimal designs are popular and some argue that they are over-used.  $c$ -optimality, on the other hand, minimizes the asymptotic variance of a user-selected function in terms of  $\theta$ , say  $c(\theta)$ . For instance, we may want to estimate the turning point of  $\eta(t, \theta)$  or its maximum, or simply a linear combination of the model parameters. In the latter case, the function we want to estimate is  $c^T \theta$  and the vector  $c$  is

user-selected. Our goal is to find a design that minimizes the asymptotic variance of the estimated  $c^T\theta$  and the optimality criterion is  $\Phi_c[M(\xi)] = c^T M^{-1}(\xi)c$ .

It is known that these criteria are convex and differentiable functions of the information matrices [23]. Designs that minimize the above two criteria are called a  $D$ - and  $c$ -optimal designs, respectively. To verify whether a given design is optimal, one resorts to standard convex analysis arguments via directional derivatives and show that when the convex criterion  $\Phi$  is differentiable, the design  $\xi^*$  is  $\Phi$ -optimal if and only if,

$$\Psi(t, \xi^*) = f^T(t) \nabla \Phi(M(\xi^*)) f(t) - \text{tr} M(\xi^*) \nabla \Phi(M(\xi^*)) \geq 0, \quad t \in \mathcal{T},$$

with equality at least at the support points of  $\xi^*$ . Here  $\nabla \Phi(\xi)$  denotes the gradient of  $\Phi(\xi)$  and the above inequality is referred to, in optimal design literature, as an Equivalence Theorem. Details are available in design monographs by Fedorov [5] and Silvey [23], for example, where algorithms and convergence issues for generating  $D$ - and  $c$ -optimal designs are also discussed.

Following convention, the worth of a design  $\xi$  is measured by its efficiency. If the criterion is  $\Phi(M(\xi))$  and  $\xi$  is an arbitrary design, the  $\Phi$ -efficiency of  $\xi$  is defined by  $\text{eff}_\Phi(M(\xi)) = \Phi(M(\xi^*)) / \Phi(M(\xi))$ . For a statistical interpretation, this definition requires a positively homogeneous criterion, that is  $\Phi(\lambda M) = \Phi(M) / \lambda$ ,  $\lambda > 0$ . The efficiency is usually multiplied by 100 and reported in percentage. If the value is 50%, this means that the design  $\xi$  has to be replicated twice to perform as well as the optimal design  $\xi^*$ .

### 3. Optimal experimental designs

We now provide two illustrative examples how optimal sampling times for estimating model parameters in the model (1) used by Jolicoeur and Pontier [9] to study growth and decline rate for *Paramecium caudatum* and *Meromyza variegata*. The first concerns an experimental population in a mixed culture where that species competed unsuccessfully with another species and the second concerns the seasonal increase and decrease of the numbers of male flies of the species. Further details of the study are given in Gause [7] and Hughes [8]. These two populations have very different behavior. The fitted model (1) for the two populations are, respectively

$$\hat{Y}_t = \frac{1}{0.010e^{t/7.192} + 0.538e^{-t/0.862}}$$

and

$$\hat{Y}_t = \frac{1}{0.009e^{t/21.419} + 1.740e^{-t/3.280}}.$$

Figure 2 shows that there is a growth at the beginning from zero to some maximum point when the size of the population starts to decline to zero asymptotically. The rates of increase and decrease in the curves depend on the population, and for these two populations, both growth rates exhibit a steep increase to their peaks and followed by a gradual decline.

Model (1) is nonlinear in the four unknown parameters  $(C_1, C_2, D_1, D_2)$ . Frequently, as in this case, we reparametrize the model to one with fewer nonlinear parameters; in this case, we have  $B = D_1$ ,  $D = D_2$  and  $C = C_2/C_1$ , and one linear parameter  $A = 1/C_1$ :

$$E(Y_t) = \frac{A}{e^{t/B} + Ce^{-t/D}}. \quad (3)$$

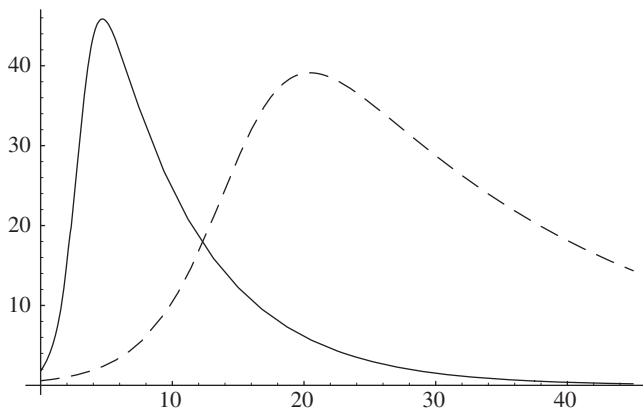


Figure 2. Comparison of the estimated models for the populations of *Paramecium* (—) and *Meromyza* (- -).

Two important features of model (1) are the maximum response (population size)  $Y(t_{\max})$  and the point where this maximum is reached,  $t_{\max}$ . A direct calculation shows

$$t_{\max} = \frac{D_1 D_2}{D_1 + D_2} \ln \frac{C_2 D_1}{C_1 D_2}, \quad (4)$$

$$Y(t_{\max}) = \frac{D_2}{C_2(D_1 + D_2)} \left( \frac{C_2 D_1}{C_1 D_2} \right)^{D_1/D_1 + D_2} \quad (5)$$

The expression of  $t_{\max}$  does not depend on  $A$  and depends on  $D_1$ ,  $D_2$ , and the ratio  $C = C_2/C_1$ . In terms of the new parameters, we have

$$t_{\max} = \frac{BD}{B+D} \ln \frac{BC}{D} \quad \text{and} \quad Y(t_{\max}) = \frac{AD}{C(B+D)} \left( \frac{BC}{D} \right)^{B/B+D}.$$

Both these quantities are functions of the model parameters and the design problem is to find a design that minimizes the asymptotic variance of each of the estimates.

To determine optimal designs for model (3), we first calculate the Fisher information matrix defined in Equation (2). This Fisher information matrix uses the linearized model obtained from the first-order Taylor expansion. For the re-parametrized model (3), the parameter vector is  $\theta = (A, B, C, D)^T$  and the vector of partial derivatives with respect to the parameters is

$$f(t, \theta) = \frac{1}{e^{t/B} + Ce^{-t/D}} \left( 1, \frac{Ae^{t/B}t}{B^2(e^{t/B} + Ce^{-t/D})}, \frac{-Ae^{-t/D}}{e^{t/B} + Ce^{-t/D}}, \frac{-ACe^{-t/D}t}{D^2(e^{t/B} + Ce^{-t/D})} \right)^T.$$

Because of the complexity of this function, we applied an algorithm to find the optimal design numerically. The optimality or proximity of the generated design to the optimal (without knowing its optimum) is then verified using an equivalence theorem.

### 3.1 D-optimal designs

Let  $C_1^{(0)}$ ,  $C_2^{(0)}$ ,  $D_1^{(0)}$  and  $D_2^{(0)}$  be nominal values of the parameters for the original model (1) and  $A^{(0)}$ ,  $B^{(0)}$ ,  $C^{(0)}$  and  $D^{(0)}$  for the re-parametrized model (3). Locally  $D$ -optimal designs were

computed for the re-parametrized model using the software *Gauss* for a wide range of the nominal values. The main general characteristics of the locally  $D$ -optimal designs are:

- (1) The  $D$ -optimal design has four equally weighted support points.
- (2) The  $D$ -optimal design depends on  $B^{(0)}$ ,  $C^{(0)}$  and  $D^{(0)}$  but not on  $A^{(0)}$ .
- (3) The larger the value of either  $B^{(0)}$  or  $C^{(0)}$  is, the larger the  $D$ -optimal design support points are.
- (4) The smaller the value of either  $B^{(0)}$  or  $C^{(0)}$  is, the more likely 0 is a support point of the  $D$ -optimal design.
- (5) The four support points increase as  $D^{(0)}$  increases. Beyond a certain value of  $D^{(0)}$ , which depends on  $B^{(0)}$  and  $C^{(0)}$ , 0 is again a support point of the  $D$ -optimal design and the other three points also increase but more slowly.

Some of these characteristics can be proved to be true and actually applies more generally to many other models using algebra. For example, Property (1) asserts locally  $D$ -optimal designs are equally weighted, meaning that the design requires an equal number of observations at each time point. The mathematical justification for Property (1) relies on the arithmetic-geometric inequality and is given in [23]. These properties similarly apply to the optimal design for the original model. The  $D$ -optimal design for the original model is the same as the one found here except for a constant ratio  $C_2^{(0)}/C_1^{(0)}$  because the  $D$ -optimal design for the re-parametrized model does not depend on  $A^{(0)}$ . In both cases, the locally  $D$ -optimal designs are equally weighted at four points; for the *P. caudatum* example, these points are at 1.9, 3.5, 5.7 and 13.2 and for the *M. variegata* example, the four points are at 10.58, 16.40, 23.87 and 46.55. We alert the reader that these are locally  $D$ -optimal designs and so these optimal sampling time points do depend on the nominal values assumed. If the nominal values for any of the parameters  $A^{(0)}$ ,  $B^{(0)}$ ,  $C^{(0)}$  and  $D^{(0)}$  change, the optimal sampling time points may also change.

It is interesting to set certain parameters in the model to selected values as described in Section 1 and consider the simpler and commonly used submodels in practice. The locally  $D$ -optimal designs for these models are reported below:

- Exponential growth model: If  $t \in [0, b]$ , the  $D$ -optimal design has the same weight at  $b - D^{(0)}$  and  $b$  when  $b - D^{(0)} \geq 0$  and at points 0 and  $b$  if  $b - D^{(0)} < 0$ .
- Logistic growth model: The  $D$ -optimal design is equally supported at three points. On the design interval  $[0, b]$ , the points may be either 0,  $t_2$  and  $b$  or  $t_1$ ,  $t_2$  and  $b$  depending on the nominal values of  $C^{(0)}$  and  $D^{(0)}$ . The latter case occurs when  $C^{(0)}$  is much larger than  $D^{(0)}$  and the interior point is a solution of a complex equation derived from an equivalence theorem. For example, in the former case, the interior support point  $t_2$  is the solution of the equation:

$$b[e^{t_2/D^{(0)}} + C^{(0)}(-1 + 2e^{t_2/D^{(0)}})] = (-1 + e^{b/D^{(0)}})[e^{t_2/D^{(0)}}(D^{(0)} - t_2) + C^{(0)}(D^{(0)} + t_2)].$$

Conradie [3] used this model to study the changes in the population of *Leysera gnaphalodes*. The response variable  $Y_t$  is the frequency of the population at time  $t$  in years,  $A$  is the ecological capacity and  $1/D$  is the maximal annual rate growth in a non-restricted population. When  $t = 0$ ,  $Y_0 = A/(1 + C)$  and we have  $C = A/Y_0 - 1$ . Thus  $C$  is determined by the initial size of the population. The fitted model is  $\hat{Y} = 89.30/(1 + 656.78e^{-3.40t})$ ,  $t \in [0, 7]$  and the  $D$ -optimal design is equally supported at 1.6, 2.2 and 7.

- Exponential decline model: If  $t \in [0, b]$  the  $D$ -optimal design is equally supported at either 0 and  $B^{(0)}$  or at 0 and  $b$ ; the former occurs if  $b - B^{(0)} \geq 0$  and the latter occurs if  $b - B^{(0)} < 0$  [21]. Dette *et al.* [4] provided  $D$ -optimal designs for other forms of the design intervals.



- Logistic decline model: On the interval  $[0, b]$ , the  $D$ -optimal design is equally supported at three points and two of them are at the extreme ends of the design interval. This model was used by Straetemans *et al.* [25] in a drug combination experiment as part of an oncology study to examine the inhibition effect of a combination of two compounds. The response variable  $Y$  is the measured radioactivity and the independent variable was the logarithm of the concentration,  $t$ . Data from the oncology study were used and the fitted homoscedastic model is approximately  $\hat{Y}_t = 7080/(1.9 + e^{t/1.12})$ ,  $t \in [-2.5, 2.1]$ . Straetemans *et al.* used a nine-point design. The locally  $D$ -optimal design obtained numerically is equally supported at the points  $-2.5$ ,  $-0.1$  and  $2.1$ . The  $D$ -efficiency of the design used in Straetemans *et al.* is about 77%.

### 3.2 Locally $c$ -optimal designs

Locally  $c$ -optimal designs are used to estimate a function of the model parameter. This function depends on the object of interest and can be (i) simply a linear function or (ii) a nonlinear function of the model parameters. We discuss these two cases separately. A simple example for case (i) is when each parameter of the original model represents a characteristic of the population and so it makes sense to find an optimal design to estimate the particular parameter as precise as possible. In this case, we want to estimate  $c^T\theta$  and  $c$  is the zero vector but with one of its entries equal to one. The  $c$ -optimal design minimizes the variance of the estimate over all designs on the design interval. Using nominal values from the two examples in [9], we determine locally  $c$ -optimal designs for different vectors  $c$ . Table 1 shows that all the locally  $c$ -optimal designs for each example are supported at the same points with different weights. In both examples, the distribution of the design weights are similar even though the optimal design points are quite different. A reason for this is the different nominal values used in each example.

Case (ii) arises when we are interested to estimate the maximum response  $Y(t_{\max})$  or the point  $t_{\max}$  where the maximum is reached. In such cases, we want a design that minimizes the asymptotic variance of the estimated quantity of interest. Specifically, if  $c(\theta)$  is the function of interest, we want to find a design  $\xi^*$  that minimizes  $\partial c(\theta)/\partial \theta M(\xi, \theta)^{-1} \partial c(\theta)/\partial \theta^T$  when  $\theta$  is set equal to the nominal value.

The above numerically optimal designs were found using a computer algorithm. Details of the algorithm can be found in standard design monograph such as Fedorov [5] and Silvey [23]. For completeness, we provide the justifications of the optimal designs in the appendix. The optimal designs for estimating the maximum response are one-point designs; at 4.7 for the *P. caudatum* example and 20.5 for the *M. variegata* example. For both examples, the  $c$ -optimal designs for estimating  $t_{\max}$ , the point where the maximum is reached have three optimal sampling time points. For the *P. caudatum* example, the points are 1.32, 3.45 and 6.15 with corresponding weights 0.19, 0.4 and 0.41, respectively. For the *M. variegata* example, the points are 21, 52 and 54 with corresponding weights 0.17, 0.41 and 0.42, respectively.

## 4. Discussion

We now study robustness properties of optimal designs for model (1). It is obvious, all optimal designs are constructed from an assumed model and the important question to ask is how sensitive the optimal designs are to model mis-specification. To investigate this issue, we evaluate the efficiencies of the locally  $D$ -optimal designs under three common and important situations: (1) when there are mis-specifications in nominal values for the model parameters, (2) the model is mis-specified and (3) under a different optimality criterion or goal. These are useful studies to undertake to gain further insights into the optimal design before implementation [18,28]. We also briefly discuss alternative and more complicated designs for the model.

Table 2.  $D$ -efficiencies of  $D$ -optimal design for the *P. caudatum* example when the nominal values  $B^{(0)}$ ,  $C^{(0)}$  and  $D^{(0)}$  change.

$C^{(0)}$	28	28	28	53	53	53	78	78	78
$D^{(0)}$	0.66	0.86	1.06	0.66	0.86	1.06	0.66	0.86	1.06
$B^{(0)} = 4.19$	0.37	0.65	0.82	0.54	0.82	0.91	0.64	0.88	0.89
$B^{(0)} = 7.19$	0.62	0.92	0.98	0.82	1	0.90	0.91	0.97	0.78
$B^{(0)} = 10.19$	0.70	0.94	0.91	0.88	0.95	0.77	0.94	0.88	0.64

#### 4.1 Efficiencies under mis-specification of the model parameters

The nominal values for the model parameters used in the *P. caudatum* data from Jolicoeur and Pontier [9] were  $A^{(0)} = 98.56$ ,  $B^{(0)} = 7.19$ ,  $C^{(0)} = 53.00$  and  $D^{(0)} = 0.86$ . Nominal values are usually available from pilot studies or similar studies. The locally  $D$ -optimal design depends on  $B^{(0)}$ ,  $C^{(0)}$  and  $D^{(0)}$  and it is interesting to investigate whether there is a substantial loss in efficiency when these nominal values are mis-specified. To do this, we first find the numerically local  $D$ -optimal designs for a set of different nominal values, say,  $B^{(0)} = 7.19 \pm 3$ ,  $C^{(0)} = 53.00 \pm 25$  and  $D^{(0)} = 0.86 \pm 0.2$ . These ranges were based on the standard errors for the estimated parameters reported in [9], but other ranges can be used as well. We then calculated the  $D$ -efficiencies of these  $D$ -optimal design relative to the one with the original nominal values  $B^{(0)} = 7.19$ ,  $C^{(0)} = 53.00$  and  $D^{(0)} = 0.86$ . Table 2 displays these efficiencies.

Table 2 shows efficiencies of the locally  $D$ -optimal design under the nominal mis-specification. The main and clear message from the table is that the optimal design performs in a non-systematic way when nominal values of the full set of parameters are varied. As an example, we observe that when  $C^{(0)} = 78$  and  $D^{(0)} = 0.86$ , the efficiencies vary from 88% to 97% and  $B^{(0)}$  varies from 4.19 to 10.19, suggesting that the optimal design is quite robust when  $B^{(0)}$  is mis-specified at this specific set of nominal values for  $C^{(0)}$  and  $D^{(0)}$ . When this set of nominal values is  $C^{(0)} = 28$  and  $D^{(0)} = 0.66$ , mis-specification in  $B^{(0)}$  will cause the  $D$ -efficiencies of the  $D$ -optimal design for model (1) to vary from 37% to 70%. For this example, the lowest efficiency of the locally  $D$ -optimal design is 37%, suggesting that mis-specification in nominal values can result in an unacceptable loss in efficiency. It is therefore prudent to always perform a robustness study before the design is implemented.

We computed the efficiencies of the locally  $D$ - and  $c$ -optimal designs for the *P. caudatum* example when they are used to estimate the model for the *M. variegata* example, and vice versa. All the efficiencies are almost zero. This is not surprising because the nominal values in both examples are quite different and nominal values for the model parameters can have a substantial impact on the resulting design.

#### 4.2 Efficiencies under different a model assumption

We also computed  $D$ -efficiencies of the  $D$ -optimal designs for model (1) when they are used for the submodels. In order to obtain  $D$ -optimal designs for the submodels, we assume the time intervals over which we can sample are  $[0, 13.2]$  and  $[0, 46.55]$ , respectively for the *Paramecium* and *Meromyza* populations. Table 3 shows the efficiencies of the designs when a submodel holds. It is remarkable that the  $D$ -optimal designs for model (1) have very low efficiencies for the growth exponential model in both populations. The practical implication here is that optimal design can be very sensitive to model specification, and generally does not perform well under another model, even within the same class of models as we see here.

Table 3.  $D$ -efficiency of the  $D$ -optimal design for model (1) when a submodel is assumed.

	Growth exp.	Growth log.	Decline exp.	Decline log.
$D$ -optimal ( <i>Paramecium</i> )	0.00	0.82	0.45	0.58
$D$ -optimal ( <i>Meromyza</i> )	0.01	0.82	0.29	0.56

Table 4. Efficiencies of the optimal designs for the *P. caudatum* (*M. variegata*) population.

	$D$ -eff	$c_{C_1}$ -eff	$c_{C_2}$ -eff	$c_{D_1}$ -eff	$c_{D_2}$ -eff
$D$ -optimal	1 (1)	0.60 (0.60)	0.48 (0.54)	0.46 (0.47)	0.67 (0.67)
$c_{C_1}$ -optimal	0.60 (0.61)	1 (1)	0.12 (0.16)	0.83 (0.88)	0.22 (0.23)
$c_{C_2}$ -optimal	0.50 (0.59)	0.15 (0.20)	1 (1)	0.09 (0.14)	0.77 (0.89)
$c_{D_1}$ -optimal	0.42 (0.45)	0.80 (0.86)	0.07 (0.10)	1 (1)	0.12 (0.14)
$c_{D_2}$ -optimal	0.76 (0.76)	0.31 (0.32)	0.81 (0.90)	0.20 (0.22)	1 (1)

4.3 Efficiencies under another criterion

We now compare the performance of the  $D$ - and  $c$ -optimal designs under a variation of optimality criteria (Table 4). The behavior of the efficiencies is very similar in both examples even though the assumed nominal values are quite different. The largest efficiencies correspond to  $c_{C_i}$ -optimal designs when they are used to estimate  $D_i$ ,  $i = 1, 2$ . The lowest efficiencies are those corresponding to the  $c$ -optimal designs for estimating  $D_i$  (respectively,  $C_i$ ) when they are used to estimate  $D_j$  (respectively,  $C_j$ ),  $i, j = 1, 2, i \neq j$ .

The locally  $c$ -optimal designs for estimating the maximum response and time to maximal response can be found directly using standard algorithms for generating them as described in design monographs such as Fedorov [5], Silvey [23] and Pazman [22]. For the two examples, these  $c$ -optimal designs are one-point designs but the optimal designs for estimating the time to maximum are supported at three points with uneven weights. We computed the efficiencies of the above optimal designs for estimating the maximum response and all of them are near zero. This means that the  $D$ - and  $c$ -optimal designs are not adequate at all to estimate the maximal mean response. The take-home message here is that the optimal design provides the most accurate estimate or estimates but can perform abysmally under another design criterion. This observation does not apply specifically to model (1) but more generally to all nonlinear models as well.

In summary, design issues are under-addressed in many applied fields. We illuminated these issues using a popular population growth/decline model with two real examples from the biological sciences. We constructed locally  $D$ - and  $c$ -optimal designs for estimating model parameters, singly or all the parameters in the model. We also found optimal designs for estimating the maximum population growth and the time for maximum growth in the population; interestingly, the former requires one time point and the latter requires three time points. Our robustness study shows that optimal designs are generally sensitive to model mis-specification, optimality criteria and also perform unpredictably when the model for the mean response is changed. The practical implication is that while the optimal design is constructed from theory, it is always important to conduct a robustness study to ascertain its utility in practice. In most cases, the optimal design should be modified to meet practical concerns. One common modification is to add more points to be able to conduct a lack of fit test. The purpose of the optimal design then is to make sure the modified optimal design does not stray too far from the optimum as judged from its efficiency relative to the optimum. We advocate that this approach be applied to all general design problems.

The above optimal designs require nominal values for construction and implementation. One common way to circumvent this dependence is to use a Bayesian approach or a maximin approach.

In the former case, we assume a prior distribution on the unknown parameters is available and we construct the information matrix by averaging over the prior information. The maximin approach assumes a certain range of plausible values for each parameter is available and the maximin optimal design maximizes the minimal efficiency regardless of the true values of the model parameters as long as they are within the plausible region. Alternatively, minimax optimal design minimizes the worst possible outcome in the sense that it minimizes the maximal inefficiency that can arise from wrong nominal values. Our experience is that minimax or maximin optimal designs are more appealing to practitioners who have less difficulty in specifying a range of plausible values for each model parameter of interest. This is in contrast to asking practitioners to come up with a single best possible value for the parameter in the case of locally optimal design, or a distribution of possible values of the parameter for Bayesian optimal design. Minimax or maximin optimal design represents a compromise for the two extreme cases. However, they are usually much more difficult to find and defy analytical descriptions except for the simplest case. Some useful references for minimax optimal design are Wong [25], Müller [19], Müller and Pázman [20], King and Wong [11,12], and Tommasi and López-Fidalgo [24].

For applied statisticians or practitioners, optimal designs for specific models are not easily available. The theory behind the construction of the optimal design can be difficult and even when it is possible to do so, the task can be a time-consuming undertaking. With this in mind, the third author has constructed a web site for generating a variety of optimal designs for several commonly used models in different fields. This free interactive site is at <http://optimal-design.biostat.ucla.edu/optimal/> and it requires the user to first select a statistical model from a list of available models and also an optimality criterion. The user then supplies design parameters to generate an optimal design. The site also enables the user to evaluate efficiency of any design of interest. In particular, this option evaluates the efficiency of the modified optimal design when the user finds it necessary to make changes to the optimal design. We hope this site will facilitate and encourage practitioners to incorporate optimal design ideas in their work.

## Acknowledgement

The research work of Wong was partially supported by the NIH grant R01GM072876. This work was also sponsored by Ministerio de Educación y Ciencia and Fondos FEDER MTM2007 67211 C03-01/03 and Junta de Comunidades de Castilla-La Mancha PAI07-0019-2036.

## References

- [1] M.P.F. Berger and W.K. Wong, *Applied Optimal Designs*, John Wiley & Sons, Chichester, UK, 2007.
- [2] H. Chernoff, *Locally optimal designs for estimating parameters*, Ann. Math. Stat. 24 (1953), pp. 586–602.
- [3] J.K. Conradie, *Modelling population dynamics of *Leysera gnaphalodes* in Namaqualand*, Doctoral thesis, South Africa, 2003.
- [4] H. Dette, I. Martínez, I. Ortiz, and A. Pepelyshev, *Maximin efficient design of experiment for exponential regression models*, J. Stat. Plan. Inference 136 (2006), pp. 4397–4418.
- [5] V.V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, 1972.
- [6] K. Fujii, *Connection between growth/development and mathematical functions*, Int. J. Sport Health Sci. 4 (2006), pp. 216–232.
- [7] G.F. Gause, *The Struggle for Existence*, Baltimore, Williams and Wilkins, 1934.
- [8] R.D. Hughes, *The influence of prevailing weather on the numbers of *Meromyza variegata* Meigen (Diptera, Chloropidae) caught with a sweepnet*, J. Anim. Ecol. 24, 324–335.
- [9] P. Jolicoeur and J. Pontier, *Population growth and decline: A four-parameter generalization of the logistic curve*, J. Theor. Biol. 141 (1989), pp. 563–571.
- [10] P. Jolicoeur, J. Pontier, and H. Abidi, *Asymptotic models for the longitudinal growth of human stature*, Am. J. Human Biol. 4 (1992), pp. 461–468.
- [11] P. Jolicoeur, J. Pontier, M.O. Pernin, and M. Sempé, *A lifetime asymptotic growth curve for human height*, Biometrics 44 (1988), pp. 995–1003.

- [12] A. Karkach, *Trajectories and models of individual growth*, Demogr. Res. 15 (2006), pp. 347–400.
- [13] J. King and W.K. Wong, *Optimal minimax designs for prediction in heteroscedastic models*, J. Stat. Plan. Inference 69 (1998), pp. 371–383.
- [14] J. King and W.K. Wong, *Minimax D-optimal designs for the logistic model*, Biometrics 56 (2000), pp. 1263–1267.
- [15] J.K. Lindsey, *Nonlinear Models in Medical Statistics*, Oxford University Press, 2001.
- [16] J. López-Fidalgo and W.K. Wong, *Design issues for the Michaelis–Menten model*, J. Theor. Biol. 215 (2002), pp. 1–11.
- [17] D.D. Martin, R.C. Hauspie, and M.B. Ranke, *Total pubertal growth and markers of puberty onset in adolescents with GHD: Comparison between mathematical growth analysis and pubertal staging methods*, Horm. Res. 63 (2005), pp. 95–101.
- [18] M. Moerbeek, *Robustness properties of A-, D-, and E-optimal designs for polynomial growth models with autocorrelated errors*, Comp. Stat. Data Anal. 48 (2005), pp. 765–778.
- [19] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, 2004.
- [20] C.H. Müller, *Maximin efficient designs for estimating nonlinear aspects in linear models*, J. Stat. Plan. Inference 44 (1995), pp. 117–132.
- [21] C.H. Müller and A. Pázman, *Applications of necessary and sufficient conditions for maximin efficient designs*, Metrika 48 (1998), pp. 1–19.
- [22] A. Pazman, *Foundations of Optimum Experimental Design*, Dordrecht, Holland, D. Reidel Publishing Company, 1986.
- [23] S.D. Silvey, *Optimal Design*, Chapman and Hall, London, 1980.
- [24] M. Stehlík, J.M. Rodríguez-Díaz, W.G. Müller, and J. López-Fidalgo, *Optimal allocation of bioassays in the case of parametrized covariance functions: An application to lung's retention of radioactive particles*, Test 17 (2008), pp. 56–68.
- [25] R. Straetmans, T. O'Brien, L. Wouters, J.V. Dun, M. Janicot, L. Bijnsens, T. Burzykowski, and M. Aerts, *Design and analysis of drug combination experiments*, Biom. J. 47 (2005), pp. 299–308.
- [26] C. Tommasi and J. López-Fidalgo, *Minimax designs for a parametrization of binary response models*, Commun. Stat. - Theory Method 33 (2004), pp. 2787–2798.
- [27] W.K. Wong, *A unified approach to the construction of mini-max designs*, Biometrika 79 (1992), pp. 611–620.
- [28] W.K. Wong, *Comparing robust properties of A, D, E and G-optimal designs*, Comput. Stat. Data Anal. 18 (1994), pp. 441–448.

## Appendix

Suppose  $Y$  is normally distributed with mean  $E(Y) = \eta(t, \theta)$ ,  $t \in \mathcal{T}$  and  $\text{Var}(Y) = 1$ , say. Let  $f(t) = \partial \eta(t, \theta) / \partial \theta$ . The Fisher information matrix at point  $t$  is defined as  $I(t, \theta) = f(t) f^T(t)$ . Suppose that there is interest in estimating a local extremum of the mean response function. The point where the extremum is reached  $v(\theta)$  solves the equation  $\partial \eta(t, \theta) / \partial t = 0$ . The goal is to find a  $c$ -optimal design to estimate the maximal response given by  $\psi(\theta) = \eta[v(\theta), \theta]$ . The vector  $c$  has to satisfy

$$\begin{aligned} c = \psi_\theta &= \frac{\partial \psi(\theta)}{\partial \theta} = \frac{\partial v(\theta)}{\partial \theta} \left( \frac{\partial \eta(t, \theta)}{\partial t} \right)_{t=v(\theta)} + \left( \frac{\partial \eta(t, \theta)}{\partial \theta} \right)_{t=v(\theta)} \\ &= \left( \frac{\partial \eta(t, \theta)}{\partial \theta} \right)_{t=v(\theta)} = f[v(\theta)]. \end{aligned}$$

Using a celebrated geometric argument from the Elfving theorem, it follows that if  $f[v(\theta)]$  is a point on the boundary of the convex hull of the Elfving set, the  $c$ -optimal design is a one-point design concentrated at  $v(\theta)$ . The optimal design for estimating  $v(\theta)$  is the  $c$ -optimal design where the vector  $c$  is given by

$$c = v_\theta = - \frac{(\partial^2 \eta(t, \theta) / \partial t \partial \theta)_{t=v(\theta)}}{(\partial^2 \eta(t, \theta) / \partial t^2)_{t=v(\theta)}},$$

which is obtained from the equation

$$0 = \frac{\partial}{\partial \theta} \left[ \left( \frac{\partial \eta(t, \theta)}{\partial t} \right)_{t=v(\theta)} \right] = \frac{\partial v(\theta)}{\partial \theta} \left( \frac{\partial^2 \eta(t, \theta)}{\partial t^2} \right)_{t=v(\theta)} + \left( \frac{\partial^2 \eta(t, \theta)}{\partial t \partial \theta} \right)_{t=v(\theta)}.$$